

Selective Bi-coordinate Method for Non-Stationary and Non-Smooth Resource Allocation Type Problems

I.V. KONNOV¹

*Department of System Analysis and Information Technologies,
Kazan Federal University, ul. Kremlevskaya, 18, Kazan 420008, Russia.*

Abstract

We propose a method of bi-coordinate variations for non-stationary and non-smooth optimization problems, which involve a single linear equality and box constraints. Here only approximation sequences are known instead of exact values of the cost function and parameters of the feasible set. It consists in making descent steps with respect to only two selected coordinates satisfying some special threshold rule. The method is simpler essentially than the usual gradient or dual type ones and differs from the previous known bi-coordinate ones suggested for the usual stationary and smooth problems. We establish its convergence under rather mild assumptions. Computational tests also reveal certain preferences of the proposed method over the known ones.

Key words: Optimization problems, non-stationary, non-smooth functions, linear equality constraint, box constraints, bi-coordinate variations, threshold control.

MSC codes: 90C30, 90C25, 90C06, 91B32, 68W15

¹E-mail: konn-igor@ya.ru

1 Introduction

The custom finite-dimensional optimization problem consists in finding the minimal value of some goal function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ on a feasible set D such that $D \subseteq \mathbb{R}^n$. For brevity, we write this problem as

$$\min_{x \in D} \rightarrow f(x), \quad (1)$$

its solution set is denoted by D^* and the optimal value of the function by f^* , i.e. $f^* = \inf_{x \in D} f(x)$. Many problems of optimal allocation of some resource within a composite system containing n elements can be reduced to the above format where

$$D = \{x \in X \mid \langle a, x \rangle = \beta\} \text{ and } X = [\alpha'_1, \alpha''_1] \times \dots \times [\alpha'_n, \alpha''_n], \quad (2)$$

β is a fixed number, $a = (a_1, \dots, a_n)^\top$ is a fixed vector whose coordinates are non-zero, and $\langle c, d \rangle$ denotes the usual scalar product of vectors c and d ; see e.g. [1]–[3] and references therein. Then, solution of problem (1)–(2) yields a feasible resource allocation that minimizes the total system dis-utility. Such problems arise in various fields and are investigated rather well and many rather efficient algorithms have been proposed; see e.g. [1, 3] and references therein.

However, the recent development of communication and information processing technologies reveal special features of resource allocation problems arising in these fields; see e.g. [4, 5] and references therein. Namely, they also reduce to the form (1)–(2), but have very large dimensionality, inexact and/or non-stationary parameters reflecting variability of users' behavior, and scattered necessary information. Hence, we are forced to develop methods whose iteration computation expenses and accuracy requirements are rather low and do not utilize matrix transformations at each iteration as the Newton or interior point type ones. This means that even simple coordinate-wise descent methods may appear very useful here.

Besides, the same optimization formulation is paid now a significant attention due to its various big data applications; see e.g. [6, 7] and the references therein. In fact, similar optimization problems arise in machine learning, signal, speech and image recognition and processing, and so on. These problems possess almost the same features; i.e., huge dimensionality, inexact, incomplete, and/or non-stationary data, which can be scattered within different computer networks. Moreover, they are often contain non-smooth regularization or penalty terms and rather simple constraints. As a result, even calculation of all the components of the gradient may be too hard. This fact creates certain difficulties for application of custom second and even first order optimization methods.

For this reason, we are interested in developing special low cost iterative methods, which are applicable for problems of form (1)–(2) and keep the convergence properties of the usual ones. In particular, their computational expenses per iteration should be reduced essentially.

In this paper, we intend to develop a new bi-coordinate descent method for these problems. We recall that the first bi-coordinate method for problems with one linear constraint and only lower (non-negativity) bounds for variables was proposed and substantiated in [8]. Further, these methods became very popular due to their rather good performance for data mining applications; see e.g. [9]–[11]; the detailed description of the recent versions is given e.g. in [12, 13].

However, most of these methods are based on either computation of certain marginal indices or utilization of some general error bound and Lipschitz constants for the gradient, so that finding a descent direction in these methods will require calculation of all the partial derivatives at each iteration, i.e., their iteration cost is almost the same as in the usual projection or conditional gradient methods. The other methods exploit the random coordinate choice idea, which reduces computational expenses per iteration, but may however lead to rather slow convergence.

Rather recently, a so-called selective bi-coordinate method with special threshold control and tolerances was proposed in [14] for problem (1)–(2) with $\alpha'_i = 0$ and $\alpha''_i = \infty$ for all $i = 1, \dots, n$, besides, the vector a was chosen to be the vector of units, that is, it was destined for the case of the usual simplex constraints. Its bi-coordinate descent is based on satisfying some threshold value and does not require calculation of all the partial derivatives in general. Besides, its threshold control strategy seems more flexible in comparison with the previous rules. In [15], its complexity estimate $O(1/\alpha)$, which gives the the total number of iterations for attaining the accuracy α , was established for the case where the goal function is convex and its partial bi-coordinate gradients are Lipschitz continuous. It should be noted that this method can be treated as a self-adjustment process for attaining an equilibrium state of a closed economic system; see [14, 16]. However, this method can not be applied directly to a general problem of form (1)–(2) with both upper and lower bounds having different signs, which somewhat restricts its field of applications.

The main goal of this paper is to develop a new selective bi-coordinate method, which follows the approach from [14], but becomes suitable for general non-stationary and non-smooth optimization problems of form (1)–(2). This means that only approximation sequences are known instead of exact values of the cost function and parameters of the feasible set, besides, the limit goal function f can be non-smooth. Clearly, these properties enlarges its areas of significant applications essentially. We establish its convergence and report some results of computational experiments with the new method and compare them with some related ones.

2 Basic preliminaries and assumptions

We will use the following first set of basic assumptions for problem (1)–(2).

(A1) *The feasible set D is nonempty, the set X is bounded, $a_i > 0$ for all $i \in I = \{1, \dots, n\}$, the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is locally Lipschitz on X , i.e. it is Lipschitz*

continuous in a neighborhood of any point $x \in X$.

Then problem (1)–(2) has a solution and $f^* > -\infty$. Here we notice that the positivity of a_i does not restrict the generality. In fact, if all a_i are negative, we can obtain the previous case by simple replacing β with $-\beta$. Next, we can consider a somewhat more general case where a has arbitrary non-zero coordinates. However, since the signs of lower and upper bounds are also arbitrary, we can introduce the new variables $y_i = \text{sign}(a_i)x_i$ for all $i \in I$ and insert the new bounds $\tilde{\alpha}'_i = -\alpha''_i$, $\tilde{\alpha}''_i = -\alpha'_i$ if $a_i < 0$ together with the previous ones $\tilde{\alpha}'_i = \alpha'_i$, $\tilde{\alpha}''_i = \alpha''_i$ if $a_i > 0$. In such a way, we again obtain the problem of form (1)–(2) satisfying the above assumptions; see also [13]. In the other words, we can always obtain the same sign for all the entries of a by proper changes of lower and upper bounds of variables.

We now recall some concepts and properties from Non-smooth Analysis; see [17] for more details. Since f is Lipschitz continuous in a neighborhood of $x \in X$, we can define its generalized gradient set at x :

$$\partial^\uparrow f(x) = \{g \in \mathbb{R}^n \mid \langle g, p \rangle \leq f^\uparrow(x; p)\},$$

which must be non-empty, convex and closed. Here $f^\uparrow(x; p)$ denotes the upper Clarke-Rockafellar derivative:

$$f^\uparrow(x; p) = \limsup_{y \rightarrow x, \alpha \searrow 0} ((f(y + \alpha p) - f(y))/\alpha).$$

It follows that

$$f^\uparrow(x, p) = \sup_{g \in \partial^\uparrow f(x)} \langle g, p \rangle.$$

At the same time, the function f has the gradient $\nabla f(x)$ a.e. in X , furthermore, it holds that

$$\partial^\uparrow f(x) = \text{conv} \left\{ \lim_{y \rightarrow x} \nabla f(y) \mid y \in D_f, y \notin S \right\}, \quad (3)$$

where D_f denotes the set of points where f is differentiable, and S denotes an arbitrary subset of measure zero. If f is convex, then $\partial^\uparrow f(x)$ coincides with the subdifferential $\partial f(x)$ in the sense of Convex Analysis, i.e.,

$$\partial f(x) = \{g \in \mathbb{R}^n \mid f(y) - f(x) \geq \langle g, y - x \rangle \quad \forall y \in \mathbb{R}^n\}.$$

In this case, we have

$$f'(x; p) = \lim_{\alpha \rightarrow 0} ((f(x + \alpha p) - f(x))/\alpha) = \sup_{g \in \partial f(x)} \langle g, p \rangle$$

and the upper derivative coincides with the usual direction derivative:

$$f^\uparrow(x; p) = f'(x; p). \quad (4)$$

Also, if f is differentiable at x , (4) obviously holds and we have

$$f'(x; p) = \langle \nabla f(x), p \rangle \text{ and } \partial^\uparrow f(x) = \{\nabla f(x)\};$$

cf. (3).

We recall that a function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ is called

(a) *pseudo-convex* on a set X , if for each pair of points $x, y \in X$, we have

$$\varphi'(x; y - x) \geq 0 \implies \varphi(y) \geq \varphi(x);$$

(b) *semi-convex* (or *upper pseudo-convex*) if for each pair of points $x, y \in X$, we have

$$\varphi^\uparrow(x; y - x) \geq 0 \implies \varphi(y) \geq \varphi(x);$$

see [18] and also [19]. In case (4), these concepts coincide, but in general (b) implies (a). Besides, the class of convex functions is strictly contained in that of pseudo-convex functions. We now recall the known optimality condition; see e.g. [17, 18] and [19].

Proposition 1 (a) *Each solution of problem (1)–(2) is a solution of the variational inequality (VI for short): Find a point $x^* \in D$ such that*

$$\exists g^* \in \partial^\uparrow f(x^*), \quad \langle g^*, x - x^* \rangle \geq 0 \quad \forall x \in D. \quad (5)$$

(b) *If f is semi-convex, then each solution of VI (5) solves problem (1)–(2).*

Solutions of VI (5) are called stationary points of (1). It will be suitable to specialize optimality conditions for the constraints in (2).

Proposition 2 *A point x^* is a solution of VI (5), (2) if and only if it satisfies each of the following equivalent conditions:*

$$x^* \in D, \exists g^* \in \partial^\uparrow f(x^*), \exists \lambda, \langle g^* - \lambda a, x - x^* \rangle \geq 0 \quad \forall x \in X; \quad (6)$$

$$x^* \in D, \exists g^* \in \partial^\uparrow f(x^*), \exists \lambda, (1/a_i)g_i^* \begin{cases} \geq \lambda & \text{if } x_i^* = \alpha'_i, \\ = \lambda & \text{if } x_i^* \in (\alpha'_i, \alpha''_i), \\ \leq \lambda & \text{if } x_i^* = \alpha''_i, \end{cases} \quad \text{for } i \in I; \quad (7)$$

$$x^* \in D, \exists g^* \in \partial^\uparrow f(x^*), \quad \forall i, j \in I, i \neq j, \\ (1/a_i)g_i^* > (1/a_j)g_j^* \implies x_i^* = \alpha'_i \text{ or } x_j^* = \alpha''_j; \quad (8)$$

$$x^* \in D, \exists g^* \in \partial^\uparrow f(x^*), \quad \forall i, j \in I, i \neq j, \\ x_i^* \in (\alpha'_i, \alpha''_i], x_j^* \in [\alpha'_j, \alpha''_j) \implies (1/a_i)g_i^* \leq (1/a_j)g_j^*. \quad (9)$$

Proof. In fact, equivalence of (5) and (6) follows from the usual optimality conditions for VIs; see e.g. [19, Theorem 12.3]. The equivalence of (6) and (7) is obvious; see e.g. [2, Proposition 7.2].

For brevity, set $h_i = (1/a_i)g_i^*$ and $h_j = (1/a_j)g_j^*$. Let now a point $x^* \in D$ satisfy (7). If there exist $i, j \in I, i \neq j$ such that $h_i > h_j, x_i^* > \alpha'_i$, and $x_j^* < \alpha''_j$, then $h_i \leq \lambda$

and $h_j \geq \lambda$, which is a contradiction. Hence, (7) implies (8). Clearly, (8) implies (9). Let now a point $x^* \in D$ satisfy (9). Define the index sets: $I_- = \{i \in I \mid x_i^* = \alpha_i'\}$, $I_0 = \{i \in I \mid x_i^* \in (\alpha_i', \alpha_i'')\}$, $I_+ = \{i \in I \mid x_i^* = \alpha_i''\}$.

If $I_0 \neq \emptyset$, set $\lambda = h_s$ for some $s \in I_0$. Then $\lambda = h_i$ for any $i \in I_0$, $\lambda \leq h_i$ for any $i \in I_-$, and $\lambda \geq h_i$ for any $i \in I_+$ due to (9), hence (7) holds.

Let now $I_0 = \emptyset$. Set $\tau_1 = \max_{i \in I_+} h_i$ and $\tau_2 = \min_{i \in I_-} h_i$, then (9) gives $\tau_1 \leq \tau_2$. Take any number $\lambda \in [\tau_1, \tau_2]$, then, by definition, $\lambda \leq h_i$ for any $i \in I_-$, and $\lambda \geq h_i$ for any $i \in I_+$, which also yields (7). \square

We intend to consider the case of the non-stationary optimization problem, where only sequences of approximations are known instead of the exact values. This means that we have some sequence of problems of the form:

$$\min_{x \in D_l} \rightarrow f_l(x), \quad (10)$$

where

$$D_l = \{x \in X_l \mid \langle a^l, x \rangle = \beta_l\} \text{ and } X_l = [\alpha_{1l}', \alpha_{1l}''] \times \dots \times [\alpha_{nl}', \alpha_{nl}''], \quad (11)$$

β_l is a fixed number, $a^l = (a_1^l, \dots, a_n^l)^\top$ is a fixed vector, for $l = 0, 1, 2, \dots$. The basic approximation assumptions are the following.

(A2) For each $l = 0, 1, 2, \dots$, the set D_l is nonempty, $a_i^l > 0$ and $-\infty < \alpha_{il}' < \alpha_{il}'' < +\infty$ for $i = 1, \dots, n$,

$$\begin{aligned} \lim_{l \rightarrow \infty} \alpha_{il}' &= \alpha_i', \lim_{l \rightarrow \infty} \alpha_{il}'' = \alpha_i'' \text{ for } i = 1, \dots, n; \\ \lim_{l \rightarrow \infty} a^l &= a, \lim_{l \rightarrow \infty} \beta_l = \beta. \end{aligned}$$

(A3) Each function $f_l : X_l \rightarrow \mathbb{R}$ is smooth, the relations $\{y^l\} \rightarrow \bar{y}$ and $y^l \in D_l$ imply $\{f_l'(y^l)\} \rightarrow \bar{g} \in \partial^\dagger f(\bar{y})$.

Assumption **(A2)** means that the sequence of the sets $\{D_l\}$ converges to the limit feasible set D , whereas **(A3)** determines some convergence property of the sequence of the differentiable functions f_l to the non-differentiable function f . These assumptions do not seem too restrictive because they do not include evaluation and concordance of deviations. In fact, **(A3)** may be invoked by several circumstances. Firstly, the limit function f of the initial problem may be smooth, and we replace it with more suitable approximations (say, if f is only convex, we can take f_l strongly convex) or f_l remains smooth despite the inexact calculation of coefficients of f . Next, if f is non-smooth, we can replace it with its smooth approximations. This technique is well known; see e.g. [20]–[23]. Since f is locally Lipschitz, it is easy to find such an approximation satisfying **(A3)**; see [20, 22]. There are simple examples for most popular non-smooth functions. For instance, we can replace $|\tau|$ with $\mu_1(\tau, \varepsilon) = \sqrt{\tau^2 + \varepsilon}$ or

$$\mu_2(\tau, \varepsilon) = \begin{cases} \tau^2/2 & \text{if } |\tau| \leq \varepsilon, \\ \varepsilon\tau - \varepsilon^2/2 & \text{if } |\tau| > \varepsilon; \end{cases}$$

where $\varepsilon > 0$ is an approximation parameter. Nevertheless, we can take into account all the opportunities mentioned above for approximating the goal function f in order to enhance the solution method performance.

3 Some examples of applications

We intend now to give some examples of applied problems which reduce to an optimization problem of form (1)–(2) and satisfy the above assumptions.

3.1 Data classification problems

One of the most popular approaches to data classification is support vector machine techniques; see e.g. [6, 24]. The simplest linear support vector machine problem for binary data classification consists in finding a hyperplane separating two collections of known points $b^i \in \mathbb{R}^m$, $i = 1, \dots, l$ attributed to some observations with different labels $\gamma_i \in \{-1, +1\}$, $i = 1, \dots, l$, where m is the number of features. That is, the distance between the hyperplane and each collection should be as long as possible. This separation of the feature space enables us to classify new data points. However, this requirement appears too strong for real problems where the so-called soft margin approach, which minimizes the penalties for mis-classification, is utilized. This problem can be formulated as the optimization problem

$$\min_{w \in \mathbb{R}^n} \rightarrow (1/p) \|w\|_p^p + C \sum_{i=1}^l L(\langle w, b^i \rangle - \beta; \gamma_i)^q,$$

where L is a loss function and $C > 0$ is a penalty parameter. The custom choice is $L(z; y) = \max\{0; 1 - yz\}$ whereas p and q are either 1 or 2. The more usual 2-norm provides so useful properties as smoothness of cost functions and uniqueness of solution, but the 1-norm approach (see [25, 26]) is also very popular since it yields sparsity, i.e. only few solution components appear non-zero. Due to very large dimensionality of the feature space, this property is valuable. If we take $p = q = 1$, we can rewrite this problem as

$$\min_{w, \xi} \rightarrow \sum_{j=1}^m |w_j| + C \sum_{i=1}^l \xi_i,$$

subject to

$$1 + \gamma_i(\beta - \langle w, b^i \rangle) \leq \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, l;$$

or in the equivalent linear programming format:

$$\min_{u, v, \xi} \rightarrow \sum_{j=1}^m (u_j + v_j) + C \sum_{i=1}^l \xi_i, \tag{12}$$

subject to

$$1 + \gamma_i \left\{ \beta - \sum_{j=1}^m (u_j - v_j) b_j^i \right\} \leq \xi_i, \quad i = 1, \dots, l;$$

$$u_j \geq 0, v_j \geq 0, \quad j = 1, \dots, m; \quad \xi_i \geq 0, \quad i = 1, \dots, l;$$

where $w_j = u_j - v_j$, $u_j \geq 0$, $v_j \geq 0$, and $|w_j| = u_j + v_j$. We can write now its dual formulation:

$$\max_y \rightarrow \sum_{i=1}^l y_i, \tag{13}$$

subject to

$$-1 \leq \sum_{i=1}^l a_{ij} y_i \leq 1, \quad i = 1, \dots, l;$$

$$\sum_{i=1}^l \gamma_i y_i = 0,$$

$$y_i \geq 0, \quad i = 1, \dots, l;$$

where $a_{ij} = \gamma_i b_j^i$. Utilization just (13) instead of (12) is suitable if $l \ll m$, moreover, (13) allows one to insert new data observations by simple adding new zero variables without losing the feasibility of the current point. It seems also worthwhile to replace the first series of double inequalities with the corresponding penalty:

$$\min_y \rightarrow (\tau/p) \sum_{j=1}^m \left\{ \left(\sum_{i=1}^l a_{ij} y_i - 1 \right)_+^p + \left(- \sum_{i=1}^l a_{ij} y_i - 1 \right)_+^p \right\} - \sum_{i=1}^l y_i,$$

subject to

$$\sum_{i=1}^l \gamma_i y_i = 0, \quad y_i \geq 0, \quad i = 1, \dots, l;$$

with $\tau > 0$, p is either 1 or 2, $(a)_+ = \max\{a, 0\}$. Clearly, this problem falls into format (1)–(2) and satisfies the basic assumptions of Section 2.

3.2 Portfolio selection problems

Investigations of portfolio selection problems were started in the Markowitz works [27, 28]. These problems still play significant role in various financial decisions. We recall that the problem is to distribute the investment capital among some (n) assets, i.e. one has to define the investment shares vector $x = (x_1, \dots, x_n)^\top$ such that

$$\sum_{i=1}^n x_i = 1, \quad x_i \geq 0, \quad i = 1, \dots, n;$$

the goal is to maximize the income

$$\sum_{i=1}^n \xi_i x_i,$$

where ξ_i is the precise return of the i -th asset, whose value is supposed to be random. One can thus take the mean variance and expected return values

$$V(x) = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_i x_j \text{ and } E(x) = \sum_{i=1}^n m_i x_i,$$

where c_{ij} and m_i are the corresponding covariance and mean for these random variables. In such a way this problem involves in fact two objectives since one should minimize the mean variance (risk) and maximize the expected return within a feasible investment share allocation; see also [29]. The classical scalar portfolio selection optimization problem consists in adding the inequality

$$\sum_{i=1}^n m_i x_i \geq w,$$

where w is the desired level of the expected return and in minimizing the mean variance over all the constraints. Some other formulations can be found e.g. in [29]. Note that all the coefficients of this problem are usually inexact and non-stationary. By replacing the above inequality with the corresponding penalty term in the goal function we can obtain another scalar optimization problem:

$$\min_x \rightarrow \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_i x_j + (\tau/p) \left(w - \sum_{i=1}^n m_i x_i \right)_+^p,$$

subject to

$$\sum_{i=1}^n x_i = 1, \quad x_i \geq 0, \quad i = 1, \dots, n;$$

with $\tau > 0$, p is either 1 or 2. Clearly, it falls into format (1)–(2) and satisfies the basic assumptions of Section 2.

3.3 Market equilibrium models

Let us consider a simple two-sided equilibrium market model of a homogeneous commodity, which follows those in [30, 2, 16].

The model involves m traders and l buyers of this commodity. Each i -th trader has a price function $g_i(x_i)$ and chooses his/her offer volume x_i in the capacity segment $[0, \alpha_i]$. Similarly, each j -th buyer has a price function $h_j(y_j)$ and chooses his/her bid

volume y_j in the capacity segment $[0, \beta_j]$. All the price functions are supposed to be continuous. Let b denote the value of the external excess demand. Then we can define the feasible set of offer/bid volumes

$$U = \left\{ u = (x, y) \in \mathbb{R}^{m+l} \left| \begin{array}{l} \sum_{i=1}^m x_i - \sum_{j=1}^l y_j = b, \\ x_i \in [0, \alpha_i], i = 1, \dots, m; \ y_j \in [0, \beta_j], j = 1, \dots, l. \end{array} \right. \right\} \quad (14)$$

We say that a pair (\bar{x}, \bar{y}) constitutes an *equilibrium point* if $(\bar{x}, \bar{y}) \in U$ and there exists a number $\bar{\lambda}$ such that

$$\begin{aligned} g_i(\bar{x}_i) \begin{cases} \geq \bar{\lambda} & \text{if } \bar{x}_i = 0, \\ = \bar{\lambda} & \text{if } \bar{x}_i \in (0, \alpha_i), \\ \leq \bar{\lambda} & \text{if } \bar{x}_i = \alpha_i, \end{cases} \quad h_j(\bar{y}_j) \begin{cases} \leq \bar{\lambda} & \text{if } \bar{y}_j = 0, \\ = \bar{\lambda} & \text{if } \bar{y}_j \in (0, \beta_j), \\ \geq \bar{\lambda} & \text{if } \bar{y}_j = \beta_j, \end{cases} \quad (15) \\ \text{for } i = 1, \dots, m; \quad \text{for } j = 1, \dots, l. \end{aligned}$$

Obviously, the number $\bar{\lambda}$ is the market clearing price. In fact, the minimal offer (bid) volumes correspond to traders (buyers) whose prices are greater (less) than $\bar{\lambda}$, and the maximal offer (bid) volumes correspond to traders (buyers) whose prices are less (greater) than $\bar{\lambda}$. The prices of other participants are equal to $\bar{\lambda}$ and their volumes may be arbitrary within their capacity bounds, but should be subordinated to the balance equation. In case $l = 0$ (respectively, $m = 0$), we have a market of traders (buyers) competing for shares of the indicated bid (offer) amount $|b|$.

It was shown in [30] (see also [2]), that each equilibrium point (\bar{x}, \bar{y}) is a solution of VI: Find $(\bar{x}, \bar{y}) \in U$ such that

$$\sum_{i=1}^m g_i(\bar{x}_i)(x_i - \bar{x}_i) - \sum_{j=1}^l h_j(\bar{y}_j)(y_j - \bar{y}_j) \geq 0 \quad \forall (x, y) \in U; \quad (16)$$

and conversely, if a pair (\bar{x}, \bar{y}) solves VI (16), (14), then there exists $\bar{\lambda}$ such that $(\bar{x}, \bar{y}, \bar{\lambda})$ satisfies (15). Moreover, we can define the function

$$\varphi(u) = \varphi(x, y) = \sum_{i=1}^m \mu_i(x_i) - \sum_{j=1}^l \eta_j(y_j),$$

where

$$\mu_i(x_i) = \int_0^{x_i} g_i(\tau) d\tau, \quad i = 1, \dots, m; \quad \text{and} \quad \eta_j(y_j) = \int_0^{y_j} h_j(\tau) d\tau, \quad j = 1, \dots, l.$$

Then, VI (16) is rewritten as follows:

$$\langle \varphi'(\bar{u}), u - \bar{u} \rangle \geq 0 \quad \forall u \in U$$

and it yields the optimality condition for the optimization problem:

$$\min_{u \in U} \rightarrow \varphi(u);$$

cf. (5) and (1). By setting $n = m + l$, $x_{m+j} = -y_j$ for $j = 1, \dots, l$ and proper modifying the bounds as indicated in Section 2, we obtain a particular case of problems (1)–(2) and (5), (1). The basic assumptions of Section 2 are also satisfied.

4 Method and its convergence

We now describe a two-level method of selective bi-coordinate variations (BCV for short) for optimization problem (1)–(2) and the related VI (5), (2) under assumptions (A1)–(A3). For brevity, set

$$g_{il}(x) = \frac{\partial f_l(x)}{\partial x_i} \text{ and } h_{il}(x) = g_{il}(x)/a_{il}, \text{ for } i \in I, l = 1, 2, \dots;$$

\mathbb{Z}_+ denotes the set of non-negative integers, and $\pi_V(u)$ denotes the projection of a point u on a set V . Also, given a sequence $\{\varepsilon_l\}$ and a point x , let

$$I_l^-(x) = \{i \in I \mid x_i \geq \alpha'_{il} + \varepsilon_l/a_{il}\}, \quad I_l^+(x) = \{i \in I \mid x_i \leq \alpha''_{il} - \varepsilon_l/a_{il}\}; \quad l = 1, 2, \dots$$

Method (BCV).

Initialization: Choose a point $z^0 \in D_0$, numbers $\sigma \in (0, 1)$, $\theta \in (0, 1)$, and sequences $\{\delta_l\} \searrow 0$, $\{\varepsilon_l\} \searrow 0$. Set $l = 1$.

Step 0: Set $k = 0$, $x^0 = \pi_{D_l}(z^{l-1})$.

Step 1: Choose a pair of indices $i \in I_l^-(x^k)$ and $j \in I_l^+(x^k)$ such that

$$h_{il}(x^k) - h_{jl}(x^k) \geq \delta_l, \tag{17}$$

set $\gamma_k = \min\{a_{il}(x_i^k - \alpha'_{il}), a_{jl}(\alpha''_{jl} - x_j^k)\}$, $i_k = i$, $j_k = j$ and go to Step 2. Otherwise (i.e. if (17) does not hold for all $i \in I_l^-(x^k)$ and $j \in I_l^+(x^k)$) set $z^l = x^k$, $l = l + 1$ and go to Step 0. (*Restart*)

Step 2: Set

$$d_s^k = \begin{cases} -1/a_{sl} & \text{if } s = i, \\ 1/a_{sl} & \text{if } s = j, \\ 0 & \text{otherwise;} \end{cases}$$

determine m as the smallest number in \mathbb{Z}_+ such that

$$f_l(x^k + \theta^m \gamma_k d^k) \leq f_l(x^k) + \sigma \theta^m \gamma_k \langle f'_l(x^k), d^k \rangle, \tag{18}$$

set $\lambda_k = \theta^m \gamma_k$, $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$ and go to Step 1.

Thus, the method has a two-level structure where each outer iteration (stage) l contains some number of inner iterations in k with the fixed tolerances δ_l and ε_l . Completing each stage, which is marked as restart, leads to the new approximation problem (10)–(11) with decreasing of the tolerances.

Note that $i_k \neq j_k$ due to (17), besides, $\gamma_k \geq \varepsilon_l$ and the point $x^k + \gamma_k d^k$ is always feasible. Moreover, by definition,

$$\mu_{kl} = \langle f'_l(x^k), d^k \rangle = h_{j_k, l}(x^k) - h_{i_k, l}(x^k) \leq -\delta_l < 0, \quad (19)$$

in (18). It follows that

$$f_l(x^{k+1}) \leq f_l(x^k) + \sigma \lambda_k \mu_{kl} \leq f_l(x^k) - \sigma \lambda_k \delta_l. \quad (20)$$

We first justify the linesearch.

Lemma 1 *Suppose assumptions (A2)–(A3) are fulfilled. Then the linesearch procedure in Step 2 is always finite.*

Proof. If we suppose that the linesearch procedure is infinite, then (18) does not hold and

$$(\theta^m \gamma_k)^{-1} (f_l(x^k + \theta^m \gamma_k d^k) - f_l(x^k)) > \sigma \mu_{kl},$$

for $m \rightarrow \infty$. Hence, by taking the limit we have $\mu_{kl} \geq \sigma \mu_{kl}$, hence $\mu_{kl} \geq 0$, a contradiction with $\mu_{kl} \leq -\delta_l < 0$. \square

We show that each stage is well defined.

Proposition 3 *Suppose assumptions (A2)–(A3) are fulfilled. Then the number of iterations at each stage l is finite.*

Proof. Fix any l . Since the sequence $\{x^k\}$ is contained in the bounded set D_l , it has limit points. Besides, by (20), we have

$$f_l^* = \min_{x \in D_l} f_l(x) \leq f_l(x^k)$$

and $f_l(x^{k+1}) \leq f_l(x^k) - \sigma \delta_l \lambda_k$, hence

$$\lim_{k \rightarrow \infty} \lambda_k = 0.$$

Suppose that the sequence $\{x^k\}$ is infinite. Since the set I is finite, there is a pair of indices $(i_k, j_k) = (i, j)$, which is repeated infinitely. Take the corresponding subsequence $\{k_s\}$, then $d^{k_s} = \bar{d}$, where

$$\bar{d}_t = \begin{cases} -1/a_{tl} & \text{if } t = i, \\ 1/a_{tl} & \text{if } t = j, \\ 0 & \text{otherwise.} \end{cases}$$

Without loss of generality, we can suppose that the subsequence $\{x^{k_s}\}$ converges to a point \bar{x} and due to (19) we have

$$\langle f'_l(\bar{x}), \bar{d} \rangle = \lim_{s \rightarrow \infty} \langle f'_l(x^{k_s}), \bar{d} \rangle \leq -\delta_l.$$

However, (18) does not hold for the stepsize λ_k/θ . Setting $k = k_s$ gives

$$(\lambda_{k_s}/\theta)^{-1}(f_l(x^{k_s} + (\lambda_{k_s}/\theta)\bar{d}) - f_l(x^{k_s})) > \sigma \langle f'_l(x^{k_s}), \bar{d} \rangle,$$

hence, by taking the limit $s \rightarrow \infty$ we obtain

$$\langle f'_l(\bar{x}), \bar{d} \rangle = \lim_{s \rightarrow \infty} \{(\lambda_{k_s}/\theta)^{-1}(f_l(x^{k_s} + (\lambda_{k_s}/\theta)\bar{d}) - f_l(x^{k_s}))\} \geq \sigma \langle f'_l(\bar{x}), \bar{d} \rangle,$$

i.e., $(1 - \sigma)\langle f'_l(\bar{x}), \bar{d} \rangle \geq 0$, which is a contradiction. \square

We are ready to prove convergence of the whole method.

Theorem 1 *Under assumptions (A1)–(A3) it holds that:*

- (i) *the number of changes of index k at each stage l is finite;*
- (ii) *the sequence $\{z^l\}$ generated by method (BCV) has limit points, all these limit points are solutions of VI (5), (2);*
- (iii) *if f is semi-convex, then*

$$\lim_{l \rightarrow \infty} f(z^l) = f^*; \quad (21)$$

and all the limit points of $\{z^l\}$ belong to D^ .*

Proof. Assertion (i) has been obtained in Proposition 3. Due to assumptions (A1)–(A2), the sets $\{D_l\}$ are uniformly bounded. Then the sequence $\{z^l\}$ is bounded, hence it has limit points. Take an arbitrary limit point \bar{z} of $\{z^l\}$, then $\bar{z} \in D$ due to (A2) and

$$\lim_{s \rightarrow \infty} z^{l_s} = \bar{z}.$$

Let p and q be arbitrary indices such that $\bar{z}_p \in (\alpha'_p, \alpha''_p]$ and $\bar{z}_q \in [\alpha'_q, \alpha''_q]$. Then $z^{l_s}_p \geq \alpha'_{p,l_s} + \varepsilon_{l_s}/a_{p,l_s}$ and $z^{l_s}_q \leq \alpha''_{q,l_s} - \varepsilon_{l_s}/a_{q,l_s}$, i.e., $p \in I_{l_s}^-(z^{l_s})$ and $q \in I_{l_s}^+(z^{l_s})$, for s large enough, hence

$$h_{p,l_s}(z^{l_s}) - h_{q,l_s}(z^{l_s}) \leq \delta_{l_s}$$

due to the stopping rule in Step 1. Taking here the limit $s \rightarrow \infty$ and applying (A2)–(A3), we obtain

$$(1/a_p)\bar{g}_p \leq (1/a_q)\bar{g}_q$$

for some $\bar{g} \in \partial^\uparrow f(\bar{z})$. This means that the point \bar{z} satisfies the optimality conditions (9). Due to Proposition 2, \bar{z} solves VI (5), (2) and assertion (ii) holds.

Next, if f is semi-convex, then by Proposition 1 each limit point $\bar{z} = \lim_{s \rightarrow \infty} z^{l_s}$ of $\{z^l\}$ solves problem (1)–(2) and

$$\lim_{s \rightarrow \infty} f(z^{l_s}) = f(\bar{z}) = f^*$$

due to the continuity of f . Since the subsequence $\{z^{l_s}\}$ was taken arbitrarily, this gives (21) and assertion (iii). \square

5 Modifications and applications

First of all we would like to emphasize the fact that convergence of the method (BCV) is attained without any concordance rules of approximation accuracy for the problem data, cost function, and threshold tolerances. We do not impose any condition for approximation of solution accuracy for intermediary problem (1)–(2) or the related VI (5), (2). Note that any explicit indication of this accuracy is not easy since we do not require (strong) convexity of the cost function.

The method described admits various modifications. We briefly discuss some of them now. Concerning the implementation of the method, we note that utilization of the projection onto the current feasible set D_l in Step 0 is not obligatory. The main condition is $x^0 \in D_l$, but the other additional condition $f_l(x^0) \leq f_l(z^{l-1})$ may give better performance.

We described the method with the current type Armijo linesearch procedure for more generality. However, some other stepsize rules can be applied in the method with maintaining all the results of Section 4. For instance, a linesearch procedure based on calculation of only two gradient components was proposed and substantiated in [14] for the case of convex function. Similarly, we can replace (18) with the following rule:

$$\langle f'_l(x^k + \theta^m \gamma_k d^k), d^k \rangle \leq \sigma \theta^m \gamma_k \langle f'_l(x^k), d^k \rangle,$$

or equivalently,

$$h_{j_k, l}(x^k + \theta^m \gamma_k d^k) - h_{i_k, l}(x^k + \theta^m \gamma_k d^k) \leq \sigma \theta^m \gamma_k (h_{j_k, l}(x^k) - h_{i_k, l}(x^k)).$$

Its preference stems from the fact that the vector d^k has only two non-zero coordinates.

We can even drop the linesearch and calculate the stepsize λ_k explicitly if partial gradients of the goal function are Lipschitz continuous. For the bi-coordinate methods these stepsize rules were substantiated in [13, 14]. Application of this rule to (BCV) and substantiation can be made similarly, hence we leave this part for the interested reader and refer to [13, 14] for more discussion. We only observe that the explicit stepsize rule reduces computational expenses essentially, but requires rather precise estimates of the corresponding Lipschitz constants that may create difficulties in the case of a general nonlinear cost function.

We now turn to application of the method to the market equilibrium models from Section 2. It was noticed in [14] that the selective bi-coordinate method proposed there can be treated as a decentralized dynamic exchange process for attaining equilibrium states in one-sided and two-sided markets. Each iteration is treated as a bilateral transaction for a pair of participants (economic agents) after comparison of their price difference in (17). Then the agents simultaneously change their current transaction amounts in order to keep the balance and bound constraints. The agents reduce the transaction thresholds (δ_l and ε_l) sequentially if the current values appear too big (restart); see [16] for more details and comparisons.

The results of Section 3 enlarge the field of applications of such processes essentially. In fact, it was shown in Section 2 that the corresponding market equilibrium model involves those in [14, 16] as particular cases. More precisely, both one-sided and two-sided models from [14, 16] can be written in the compact format (1)–(2) or (5), (2), which gives a simpler process definition in comparison with that in [14, Section 6]. Besides, our current formulation now handles both upper and lower bounds for variables.

Moreover, we note that after transformation of the market equilibrium model from Section 2 into format (5), (2) we can differ agents by considering signs of their volume variables. That is, $x_i > 0$ indicates the i -th offer value, whereas $x_j < 0$ indicates the j -th bid value $|x_j|$. In the models from [14, 16], both the upper and lower bounds of one agent must have the same sign, hence his/her role is fixed as either trader or buyer. However, we can now utilize different signs for upper and lower bounds of one agent in format (5), (2), say, $\alpha'_i < 0$ and $\alpha''_i > 0$. This means that the i -th agent can change his/her role in this market model. Therefore, the results of Section 3 confirm that the selective bi-coordinate method proposed can serve as a decentralized dynamic exchange process in much more complex and non-stationary market systems.

6 Computational experiments

In order to check the performance of the proposed method we carried out series of computational experiments on test problems. For comparison, we took the known conditional gradient method (CGM) [31, 32] and marginal-based bi-coordinate descent method (MBC) [8, 12] with the same Armijo linesearch procedures. We recall that the computation of the descent direction in (MBC), unlike (BCV), is based on finding the so-called most violated pair of indices. All the methods were implemented in *Delphi* with double precision arithmetic. The main goal was to compare convergence of the methods despite the smaller iteration expenses of (BCV). In all the cases, we took the accuracy $\mu = 0.1$ and the starting point $(\beta/n)e$, where e denote the vector of units in \mathbb{R}^n . We chose $\sigma = \theta = 0.5$, and the rule $\delta_{l+1} = \nu\delta_l$, $\varepsilon_{l+1} = \nu\varepsilon_l$ with $\nu = 0.5$ for (BCV). For testing, we chose problems of form (1)–(2) with $a_i = 1$, $\alpha'_i = 0$ and $\alpha''_i = 1 + (\beta/n) + 0.5 \sin(i)$ for $i = 1, \dots, n$.

In the first two series, we took the stationary problem (1)–(2) with the fixed data and smooth goal function f . We hence took the value

$$\Delta(x) = \max_{y \in D} \langle f'(x), x - y \rangle$$

as an error bound at x and write Δ_k for the accuracy $\Delta(x)$ after full k iterations. We give the total number of iterations of each method for attaining the indicated accuracy in each case, sign “-” means that the error was too big, namely, $\Delta_{500} > 1$.

In the first series, we took the quadratic cost function $f(x) = \varphi(x)$ where

$$\varphi(x) = 0.5 \langle Px, x \rangle,$$

Table 1:

	(CGM)	(BCV)	(MBC)
$\beta = 5$			
$n = 10$	66	30	$\Delta_{500} \approx 1.28$
$n = 20$	22	41	$\Delta_{500} \approx 0.99$
$n = 50$	82	96	$\Delta_{500} \approx 0.91$
$n = 100$	$\Delta_{500} \approx 0.1$	213	$\Delta_{500} \approx 1.63$
$\beta = 10$			
$n = 10$	55	40	$\Delta_{500} \approx 5.13$
$n = 20$	103	54	-
$n = 50$	90	145	-
$n = 100$	$\Delta_{500} \approx 0.48$	299	-
$\beta = 20$			
$n = 10$	$\Delta_{500} \approx 0.14$	62	-
$n = 20$	$\Delta_{500} \approx 0.23$	80	-
$n = 50$	$\Delta_{500} \approx 0.21$	191	-
$n = 100$	$\Delta_{500} \approx 1.07$	405	-

the elements of the matrix P were defined by

$$p_{ij} = \begin{cases} \sin(i) \cos(j) & \text{if } i < j, \\ \sin(j) \cos(i) & \text{if } i > j, \\ \sum_{i=1}^n |p_{ij}| + 1 & \text{if } i = j. \end{cases}$$

We varied the parameter β and dimensionality n . The results are given in Table 1.

In the second series, we took the convex cost function

$$f(x) = \varphi(x) + \psi(x),$$

where the function φ was defined as above,

$$\psi(x) = -\ln(\langle c, x \rangle + \xi),$$

the elements of the vector c are defined by

$$c_i = 2 + \sin(i) \quad \text{for } i = 1, \dots, n,$$

and $\xi = 5$. The results are given in Table 2.

In the third series, we took the non-smooth convex cost function

$$f(x) = \varphi(x) + \psi(x) + \sum_{i=1}^n |x_i|,$$

Table 2:

	(CGM)	(BCV)	(MBC)
$\beta = 5$			
$n = 10$	77	29	$\Delta_{500} \approx 1.29$
$n = 20$	30	35	-
$n = 50$	111	109	-
$n = 100$	457	240	-
$\beta = 10$			
$n = 10$	62	44	$\Delta_{500} \approx 5.14$
$n = 20$	77	53	-
$n = 50$	115	167	-
$n = 100$	$\Delta_{500} \approx 0.46$	282	-
$\beta = 20$			
$n = 10$	$\Delta_{500} \approx 0.12$	68	-
$n = 20$	$\Delta_{500} \approx 0.21$	75	-
$n = 50$	$\Delta_{500} \approx 0.24$	220	-
$n = 100$	$\Delta_{500} \approx 1.07$	350	-

where the functions φ and ψ were defined as above. We also took the fixed coefficients $a_i = 1$, $\alpha'_i = 0$ and $\alpha''_i = 1 + (\beta/n) + 0.5 \sin(i)$ for $i = 1, \dots, n$. We utilized the smooth approximations of the form

$$\phi(x, \tau) = \varphi(x) + \psi(x) + \sum_{i=1}^n \sqrt{x_i^2 + \tau^2}.$$

In other words, we replace (1)–(2) with the sequence of the following smooth optimization problems

$$\min_{x \in D} \rightarrow f_l(x),$$

where $f_l(x) = \phi(x, \tau_l)$ for some sequence $\{\tau_l\} \searrow 0$, i.e. set $D_l = D$. The main goal was to check the performance for such smooth approximations of the non-smooth initial problem. Since (MBC) appeared rather slow, we compared only (CGM) and (BCV). We used the value

$$\Delta(x, \tau) = \max_{y \in D} \langle \phi'(x, \tau), x - y \rangle$$

as an error bound at x . We stopped the calculations under the condition

$$\Delta(x, \tau) \leq \mu \text{ and } \tau \leq \mu$$

with $\mu = 0.1$. We used the rule $\tau_{l+1} = \max\{\mu, \nu \tau_l\}$ with $\nu = 0.5$, the other parameters of the methods were chosen as above. The results are given in Table 3, where $\Delta_{\tau,k}$ denotes the accuracy $\Delta(x, \tau)$ after full k iterations.

Table 3:

	(CGM)	(BCV)
$\beta = 5$		
$n = 10$	154	57
$n = 20$	43	52
$n = 50$	103	85
$n = 100$	300	234
$\beta = 10$		
$n = 10$	98	49
$n = 20$	123	52
$n = 50$	183	136
$n = 100$	$\Delta_{\tau,500} \approx 0.81, \tau = 0.8$	271
$\beta = 20$		
$n = 10$	$\Delta_{\tau,500} \approx 6.8, \tau = 6.4$	66
$n = 20$	$\Delta_{\tau,500} \approx 0.9, \tau = 0.8$	67
$n = 50$	$\Delta_{\tau,500} \approx 0.48, \tau = 0.2$	197
$n = 100$	$\Delta_{\tau,500} \approx 2.06, \tau = 1.6$	468

In all the experiments, (BCV) showed rather rapid and stable convergence. In almost all the cases, (BCV) showed better results than (CGM). At the same time, these experiments showed rather slow and instable convergence of (MBC). We also noticed that the presence of nonlinear functions or approximations of non-smooth functions had no significant influence on the convergence of (BCV).

7 Conclusions

We suggested a new method of bi-coordinate variations for non-stationary and non-smooth optimization problems, which involve two side constraints for variables and a single linear equality. Its descent direction rule selects only two coordinates for changes and enables us to avoid calculation of all the gradient components of a current smooth approximation function at each iteration. Therefore, the new method is simpler essentially than the usual gradient or dual type ones, but does not impose any concordance rules of approximation accuracy for the problem data, cost function, and threshold tolerances. We showed some its possible fields of applications. Its convergence was established under rather mild assumptions. Computational tests showed certain preferences of the proposed method over the known ones.

Acknowledgement

This work was supported by the RFBR grant, project No. 16-01-00109a and by grant No. 297689 from Academy of Finland.

References

- [1] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, London, 1989.
- [2] I.V. Konnov, *Equilibrium Models and Variational Inequalities*, Elsevier, Amsterdam, 2007.
- [3] M. Patriksson, *A survey on the continuous nonlinear resource allocation problem*, Eur. J. Oper. Res., vol. 185 (2008), pp. 1–46.
- [4] C. Courcoubetis and R. Weber, *Pricing Communication Networks: Economics, Technology and Modelling*, John Wiley and Sons, Chichester, 2003.
- [5] S. Stańczak, M. Wiczanowski, and H. Boche, *Resource Allocation in Wireless Networks. Theory and Algorithms*, Springer, Berlin, 2006.
- [6] C.J.C. Burges, *A tutorial on support vector machines for pattern recognition*, Data Mining Know. Disc., vol. 2 (1998), pp.121–167.
- [7] V. Cevher, S. Becker, and M. Schmidt, *Convex optimization for big data*, Signal Process. Magaz., vol. 31 (2014), pp.32–43.
- [8] G.M. Korpelevich, *Coordinate descent method for constrained minimization problems, linear inequalities, and matrix games*, In: Gol’shtein, E.G. (ed.) *Mathematical Methods for Solving Economic Problems*, vol.9, Nauka, Moscow, pp. 84–97, 1980. [In Russian]
- [9] J.C. Platt, *Sequential minimal optimization: a fast algorithm for training support vector machines*, In: Scholkopf, B., Burges, C.J.C., Smola, A.J. (eds.) *Advances in Kernel Methods – Support Vector Learning*, MIT Press, Cambridge, pp. 185–208, 1999.
- [10] S.S. Keerthi and E.G. Gilbert, *Convergence of a generalized SMO algorithm for SVM classifier design*, Mach. Learn., vol. 46 (2002), pp.351–360.
- [11] T. Glasmachers and C. Igel, *Maximum-gain working set selection for SVMs*, J. Mach. Learn. Res., vol.7 (2006), pp.1437–1466.

- [12] S. Lucidi, L. Palagi, A. Risi, and M. Sciandrone, *A convergent hybrid decomposition algorithm model for SVM training*, IEEE Trans. Neural Netw., vol. 20(2009), pp.1055–1060.
- [13] A. Beck, *The 2-coordinate descent method for solving double-sided simplex constrained minimization problems*, J. Optim. Theory. Appl., vol.162 (2014), pp.892–919.
- [14] I.V. Konnov, *Selective bi-coordinate variations for resource allocation type problems*, Comp. Optim. Appl., vol.64 (2016), pp. 821–842.
- [15] I.V. Konnov, *A method of bi-coordinate variations with tolerances and its convergence*, Russ. Mathem. (Iz. VUZ), vol.60 (2016), pp.68–72.
- [16] I.V. Konnov, *An alternative economic equilibrium model with different implementation mechanisms*, Adv. Model. Optim., vol. 17 (2015), pp.245–265.
- [17] F.H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.
- [18] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Contr. Optim., vol.15 (1977), pp.959–972.
- [19] I.V. Konnov, *Nonlinear Optimization and Variational Inequalities*, Kazan Univ. Press, Kazan, 2013. [In Russian]
- [20] Y.M. Ermoliev, V.I. Norkin, and R.J.B. Wets, *The minimization of semicontinuous functions: mollifier subgradient*, SIAM J. Contr. Optim., vol.33 (1995), pp.149–167.
- [21] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., vol.103 (2005), pp. 127–152.
- [22] M.-O. Czarnecki and L. Rifford, *Approximation and regularization of Lipschitz functions: Convergence of the gradients*, Trans. Amer. Math. Soc., vol.358 (2006), pp. 4467–4520.
- [23] A. Beck and M. Teboulle, *Smoothing and first order methods: A unified framework*, SIAM J. Optim., 22 (2012), pp. 557–580.
- [24] C.C. Agarval, *Data Mining*, Springer, Heidelberg, 2015.
- [25] C. Cortes and V. Vapnik, *Support vector networks*, Mach. Learn., vol.20 (1995), pp. 273–297.

- [26] P.S. Bradley and O.L. Mangasarian, *Feature selection via concave minimization and support vector machines*, In: Shavlik, J. (ed.) Machine Learning, Proc. of the 15-th Intern. Conf. (ICML 98), Morgan Kaufmann, San Francisco, pp. 82–90, 1998.
- [27] H.M. Markowitz, *Portfolio selection*, J. Finance, vol.7 (1952), pp. 77–91.
- [28] H.M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, John Wiley and Sons, New York, 1959.
- [29] R.E. Steuer, Y. Qi, and M. Hirschberger, *Multiple objectives in portfolio selection*, J. Financ. Decis. Making, vol.1 (2005), pp. 11–26.
- [30] I.V. Konnov, *On modeling of auction type markets*, Issled. Inform., vol.10 (2006), pp. 73–76. [In Russian]
- [31] M. Frank and P. Wolfe, *An algorithm for quadratic programming*, Nav. Res. Logist. Quart., vol. 3 (1956), pp.95–110.
- [32] E.S. Levitin and B.T. Polyak, *Constrained minimization methods*, USSR Comp. Maths. Math. Phys., vol. 6 (1966), pp.1–50.